AFRL-RH-WP-TR-2009-0015

An Investigation of Image Fusion Algorithms using a Visual Performance-based Image Evaluation Methodology

Kelly E. Neriani
Consortium Research Fellow Program
4214 King Street, First Floor
Alexandria VA 22302-1555

Alan R. Pinkus
Warfighter Interface Division
711th Human Performance Wing
Wright-Patterson AFB OH 45433-7022

David W. Dommett
General Dynamics-Advanced Information Engineering Systems
5200 Springfield Pike, Suite 200
Dayton OH 45431-1289

February 2009

Final Report for November 2003 to December 2008

Air Force Research Laboratory
711th Human Performance Wing
Human Effectiveness Directorate
Warfighter Interface Division
Battlespace Visualization Branch
Wright-Patterson AFB OH 45433

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th Air Base Wing Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RH-WP-TR-2009-0015 HAS BEEN REVIEWD AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR

//signed//
Alan R. Pinkus, Ph.D.
Senior Engineering Research Psychologist
Battlespace Visualization Branch

//signed//
Daniel G. Goddard
Chief, Warfighter Interface Division
Human Effectiveness Directorate
711th Human Performance Wing
Air Force Research Laboratory

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
2-Feb-2009	Final	Nov 2003 – Dec 2008

4. TITLE AND SUBTITLE
An Investigation of Image Fusion Algorithms using a Visual Performance-Based Image Evaluation Methodology

5a. CONTRACT NUMBER
In-house

5b. GRANT NUMBER
N/A

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)
Kelly E. Neriani*, Alan R. Pinkus**, David W. Dommett***

5d. PROJECT NUMBER
7184

5e. TASK NUMBER
11

5f. WORK UNIT NUMBER
71841127

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Consortium Research Fellow Program*
4214 King Street, First Floor Alexandria VA 22302-1555
General Dynamics-Advanced Information Engineering Systems***
5200 Springfield Pike, Suite 200, Dayton OH 45431-1289

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Material Command**
Air Force Research Laboratory
711 Human Performance Wing
Warfighter Interface Division
Battlespace Visualization Branch
Wright-Patterson AFB, OH 45433-7022

10. SPONSOR/MONITOR'S ACRONYM(S)
711 HPW/RHCV

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-RH-WP-TR-2009-0015

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
88 ABW PA Cleared 02/17/09; 88ABW-09-0522.

14. ABSTRACT
It is believed that the fusion of multiple different images into a single image should be of great benefit to warfighters engaged in a search task. As such, more research has focused on the improvement of algorithms designed for image fusion. Many different fusion algorithms have already been developed; however, the majority of these algorithms have not been assessed in terms of their visual performance-enhancing effects using militarily relevant scenarios. The goal of this research is to apply a visual performance-based assessment methodology to assess four algorithms that are specifically designed for fusion of multispectral digital images. The image fusion algorithms used included a Principle Component Analysis based algorithm, a Shift-invariant Wavelet transform algorithm, a Contrast-based algorithm, and pixel averaging. The methodology used has been developed to acquire objective human visual performance data as a means of evaluating the image fusion algorithms. Standard objective performance metrics (response time and error rate), were used to compare the fused images vs. two baseline conditions comprising each individual image used in the fused test images. Observers searched images for a military target hidden among foliage and then indicated in which quadrant of the screen the target was located using a spatial-forced-choice paradigm. Response time and percent correct were measured for each observer.

15. SUBJECT TERMS: Image fusion, sensor fusion, multi-sensor fusion, multispectral fusion, fusion algorithm, fusion method, multispectral image

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
Unclassified					Alan R. Pinkus
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER (*Include area code)***
U	U	U	SAR	32	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI-Std Z39-18

This page is intentionally blank.

Table of Contents

ACKNOWLEDGEMENTS v

1.0 SUMMARY 1

2.0 INTRODUCTION 1

3.0 FUSION ALGORITHMS 2

3.1 Principle Components Analysis (PCA) 2

3.2 Shift-invariant Wavelet Transform 2

3.3 Contrast-based 2

3.4 Averaging 3

4.0 METHODS 3

4.1 Observers 3

4.2 Sensors 3

4.2.1 Visible Sensor 3

4.2.2 Thermal Sensor 3

4.3 Stimuli 3

4.4 Procedure 5

5.0 RESULTS 8

6.0 DISCUSSION / CONCLUSIONS 10

7.0 REFERENCES 11

8.0 APPENDIX: ANNOTATED BIBLIOGRAPHY OF SELECT IMAGE ENHANCING ALGORITHMS 12

List of Figures

Figure 1: Sensors used in the experiment.....4
Figure 2: Target used in the experiment.....5
Figure 3: Terrain board used for scenery in the experiments.....5
Figure 4: Fixation image to ready the observer for the start of a new trial.....6
Figure 5: Image used to indicate quadrant assignment. 6
Figure 6: The six different algorithm conditions. The arrow in each image highlights the target.....7
Figure 7: Mean response time (across observers) by algorithm condition. 9
Figure 8. Mean percent error (across observers) by algorithm condition..... 9

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the excellent support provided by Sheldon E. Unger, who created the terrain board, and David W. Sivert, who collected the data, both of General Dynamics-Advanced Information Engineering Systems, Dayton OH.

This page intentionally left blank.

1.0 SUMMARY

It is believed that the fusion of multiple different images into a single image should be of great benefit to Warfighters engaged in a search task. As such, more research has focused on the improvement of algorithms designed for image fusion. Many different fusion algorithms have already been developed; however, the majority of these algorithms have not been assessed in terms of their visual performance-enhancing effects using militarily relevant scenarios. The goal of this research is to apply a visual performance-based assessment methodology to assess four algorithms that are specifically designed for fusion of multispectral digital images. The image fusion algorithms used in this study included a Principle Component Analysis (PCA) based algorithm, a Shift-invariant Wavelet transform algorithm, a Contrast-based algorithm, and the standard method of fusion, pixel averaging. The methodology used has been developed to acquire objective human visual performance data as a means of evaluating the image fusion algorithms. Standard objective performance metrics, such as response time and error rate, were used to compare the fused images versus two baseline conditions comprising each individual image used in the fused test images (an image from a visible sensor and a thermal sensor). Observers completed a visual search task using a spatial-forced-choice paradigm. Observers searched images for a target (a military vehicle) hidden among foliage and then indicated in which quadrant of the screen the target was located. Response time and percent correct were measured for each observer. Results of this study and future directions will be discussed. An annotated bibliography of select image enhancing algorithms is also included as an Appendix to this report.

2.0 INTRODUCTION

Battlefield operators are bombarded by vast amounts of visual information, often coming from multiple sources. This visual information must be constantly tracked, while operators are tasked with making critical decisions in short amounts of time. Multisensor image fusion has become a popular methodology to reduce the visual workload of an operator, while still maintaining the important details contained within a set of images. Image fusion is defined as a mathematical process of combining information from two or more images of a scene into a single composite image that is more informative for visual perception or computer processing. It is generally believed that fused images should reduce redundancy and maximize the information relevant to a particular task. By having information from two sensors combined into one image, an operator looking for a target in a scene should be able to locate a target faster and more accurately.

There are several different techniques to assess the relative improvement in image quality when an image fusion algorithm has been applied to a set of digital images. The testing of enhancing effects often consists of subjective quality assessments or measures of the ability of an automatic target detection program to find a target before and after images have been fused. It is rare to find studies that focus on the human ability to detect a target in a fused image using scenarios that are relevant for the particular application for which the enhancement is intended.

While a particular algorithm may make an image appear substantially better after enhancement, there is no indication as to whether this improvement is significant enough to improve human visual performance. Therefore, Neriani, et al.[1], developed a methodology that used a visual search task to

determine the effect, if any, that image enhancing algorithms have on improving visual performance. Since the aim of the research was to improve visual performance, an improvement in image quality was quantified as a decrease in response time when observers perform a visual search task. Specifically, the research was designed to measure performance on the militarily relevant task of searching for a target hidden among foliage. The methodology used gave a precise and useful estimate as to how much (if at all) the observers' performance improved when the target images were enhanced with each of the three Retinex algorithms that were studied.

Using the methodology developed in Neriani et al.,[1] the research discussed in this paper is focused on the assessment of four different fusion algorithms. These algorithms are described in some detail in the next section. The hypothesis of this research is that when observers are tasked with finding a heated target among other heated distractor targets, their performance in both response time and percent correct measures will be better with fused images than with unfused images from either a visible or thermal sensor. The premise is that one needs both the fine detail from the visible image and the heat information from the thermal image to adequately complete the task.

3.0 FUSION ALGORITHMS

3.1 Principle Components Analysis (PCA)

The algorithm used is taken from Kumar and Muttan.[2] This section will only briefly touch upon the fusion method. For more detail, please see the original source, where it is described in full. Principle components analysis (PCA) is a statistical technique used to decrease a large set of variables to a much smaller set that still contains most of the information that was available in the larger set. By reducing the interchannel dependencies to pull out just the principle factors (or components) of the data, it becomes much easier to analyze and interpret. The basic concept behind using a PCA method to fuse images is to find the unique information in each of the spectral bands and create a new image by fusing only that non-redundant information that contributes the most to the variation in the data set.

3.2 Shift-invariant Wavelet Transform

The algorithm used is taken from Rockinger.[3] This section will only briefly touch upon the fusion method. For more detail, please see the original source, where it is described in full. The importance of using a shift-invariant wavelet fusion method versus a standard wavelet fusion method comes into play when fusing images that have unknown object locations. In standard fusion methods, one must have fixed object locations to achieve a pleasing fused image. In the shift-invariant method, the images are decomposed such that all possible (circular) shifts of the input images are calculated. This is highly overcomplete and redundant. In the actual fusion process, the input images are decomposed into their shift-invariant wavelet representation and a fused representation is built by using an appropriate selection scheme. The paper referenced above identifies two methods of selection. The method chosen for this study is the point based *choose-max* method.

3.3 Contrast-based

The algorithm used is taken from Peli et al.[4] This section will only briefly touch upon the fusion method. For more detail, please see the original source, where it is described in full. The multispectral fusion process described in this paper has three key attributes: 1. a scale-by-scale fusion using oriented filters, 2. a fusion decision based on the contrast in each scale, and 3. a

preference for the visible band for at least larger scale sizes (to preserve shape from shading contrast). The input images are filtered at four orientations (0°, 45°, 90°, and 135°). The basic fusion process compares a calculated contrast measure for each pixel, at each scale and orientation, and selects the spectral band that should dominate in the fused image.

3.4 Averaging

The averaging fusion method simply takes the values at each pixel in both the visible and thermal image and finds the average of these two values. The fused image is then comprised of these averaged values. This method is obviously quite computationally simple, however as the literature suggests, often is not an effective fusion method for target detection.

4.0 METHODS

The goal of this research was to use a methodology that incorporates a measure of human visual performance in the assessment of the effectiveness of several multispectral fusion algorithms. A standard psychophysical task (spatial-forced-choice visual search task) was employed to measure human visual performance.

4.1 Observers

Four observers, two males and two females, participated in the experiment. The observers ranged in age from 21 to 31 years. All had normal color vision and normal or corrected to normal visual acuity. All of the observers completed at least three practice sessions on the task before the data reported were collected.

4.2 Sensors

The images to be fused came from two different sensors, each sensitive across non-overlapping wavebands. These were chosen based on their unique sensitivities, which provide distinctly different information to be fused.

4.2.1 Visible Sensor

The visible sensor used in this study was the XC-77 Monochrome Machine Vision Camera manufactured by Sony. This camera is sensitive in the range of 400 to 800 nm. This is effectively the range in which the human visual system is sensitive. Thus, the images captured by the sensor appear much as one would see with a standard digital camera. The highest resolution achievable by the camera is 768 x 493 pixels.

4.2.2 Thermal Sensor

The thermal sensor used in this study was the LTC550 fixed mount infrared camera manufactured by BAE Systems. This camera is sensitive in the range of 8 to 14 microns. The human visual system is not sensitive to wavebands in this range. Therefore, images captured by the sensor tend to appear different from what one would see given a natural view. The highest resolution achievable by the camera is 320 x 240 pixels.

4.3 Stimuli

In the experiment, each subject viewed a total of 1272 grayscale images. Of these, 1152 images consisted of a target located in a scene of trees and grass. The target was a model of a BMP-3 Armored Personnel Carrier tank (see Figure 1), placed on an artificial terrain board. Additionally, small rocks that are the same shape and size as the target (see Figure 2) were added to the terrain

board to act as distractors. Both the target and the distractor rocks were heated in an oven to be 110° when each image was taken. The terrain board was specially constructed to provide an accurate replication of the scene reflectivity that would be collected using visible and thermal cameras in a natural environment (see Figure 3). For the rest of the 1272 test images, the observers viewed 120 images of the artificial terrain board scene of trees, grass, and distractor rocks with no target as catch trials. These catch trials were used to help reduce guessing by the subject since they were told there are some trials with no target.

The test images were taken with the artificial terrain board rotated at 16 different angles (spanning from 0° - 337.5°) and at each rotation angle, the cameras were positioned on the tripod at three different tilts (cameras tilted to the left, in the center, and tilted to the right). If a target was present, the target was placed so that it clearly fell into one of the four quadrants of the screen. Additionally, to control for the effect of foliage density, the targets were always placed in an area of medium masking, which is defined as the target being located along a tree line or directly adjacent to a clump of bushes.

All 1272 images were presented using one of six different algorithm conditions. These conditions were a visible unfused image, a thermal unfused image, and the four conditions corresponding to visible and thermal images processed by each of the four fusion algorithms described above. The images were taken with the visible and thermal sensors set to be 3 meters away from the terrain board and illuminated with flood lights. The 1536 images with a target present included one trial for each combination of 16 rotation angles x 3 camera tilts x 4 quadrants x 6 algorithm conditions.

The images were taken using both the visible and infrared cameras set to their highest resolution (768 x 493, and 320 x 240 pixels, respectively) and in grayscale mode. Before the images were used in the experiment, all the visible images were then resized to be at the lower resolution of the thermal camera (320 x 240 pixels). Prior to executing any of the fusion algorithms, the visible and thermal images were registered using an in-house custom registration program such that each pixel in the visible image would perfectly correspond to its respective pixel in the thermal image.

Figure 1: Sensors used in the experiment.
The thermal sensor is on top and the visible sensor is directly under it.

Figure 2: Target used in the experiment.

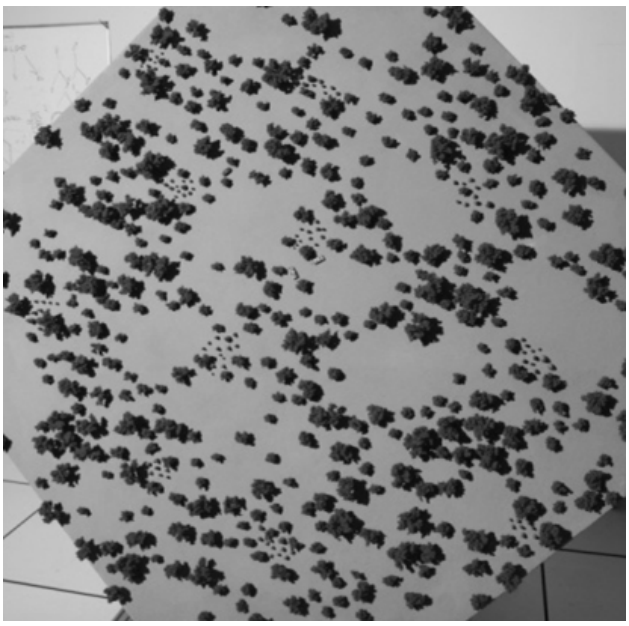


Figure 3: Terrain board used for scenery in the experiments.

4.4 Procedure

The observers viewed the images on a 17-inch Dell E173FP LCD color monitor driven by a Diamond Savage4 video card in a 700 MHz Pentium III NCS Computer placed on a desktop. The observers were seated in a comfortable chair in a dimly lit room during the experimental sessions and viewed the images from a distance of about 36 inches. Prior to the collection of data, the observers were instructed to perform a visual search on each image shown to locate a target and that once they located the target, they would be required to indicate which quadrant the target was located in. The observers were told that there would be distractors in the scene and that they must decide which item in the scene was the target. Additionally, the observers were told that some of the images shown would not have a target. Before each trial started, the observers fixated on a blank green screen with a black fixation cross in the center of it (see Figure 4). Observers pressed the spacebar on the keyboard to initiate a trial whenever they were ready. Once the spacebar was pressed, the image was displayed. The observer pressed the spacebar again as soon as they determined whether the target was present or absent and the displayed image was replaced with an image indicating the assignment of the four quadrants (see Figure 5). Then, the observer pressed

the number on the keyboard that corresponded to the quadrant the target was located in (1-4) or zero if there was no target in the image. Each trial had a time limit of 30 seconds. If the observer did not press the spacebar in 30 seconds or less once the test image was displayed, the test image was automatically removed and the quadrant screen was displayed. Observers were then forced to choose which quadrant the target was in or respond zero if they did not find the target. Percent correct and response time (measured from the first display of the image until the spacebar was pressed to indicate target presence or absence) were recorded for each trial.

Each observer had two 30 minute sessions on each of five days. During each session, the observer viewed 108 images with a target present and twelve images with no target. The presentation order across all sessions for each observer was randomized with the constraint that each of the six algorithm conditions was used for eighteen trials within each session.

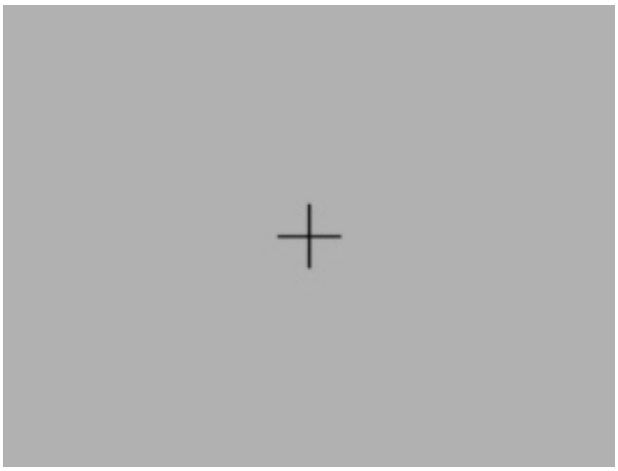


Figure 4: Fixation image to ready the observer for the start of a new trial.

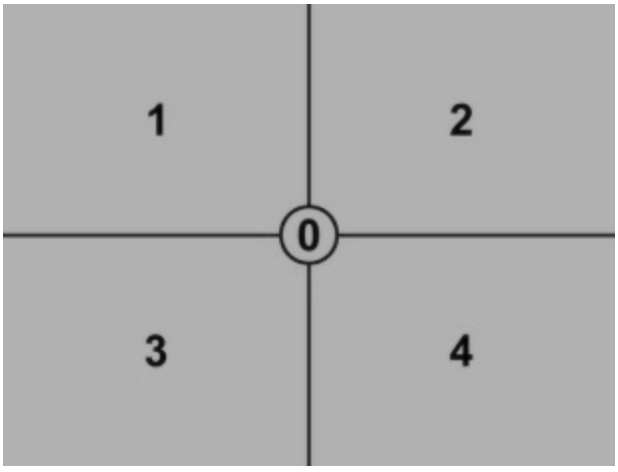
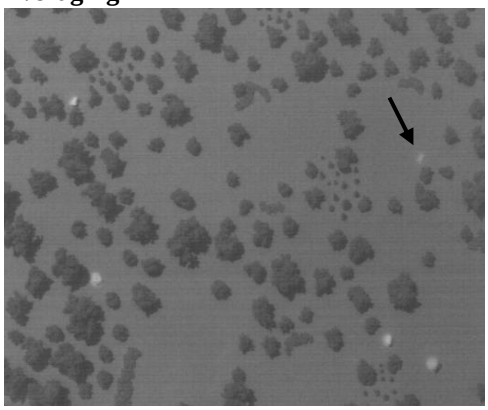


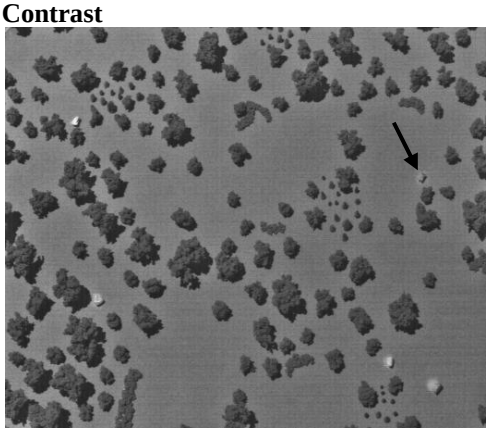
Figure 5: Image used to indicate quadrant assignment.

Figure 6 shows the six different algorithm conditions applied to one scene (be aware of differences between images displayed on the CRT and images displayed on paper). The arrow in each image highlights the target, which is placed in the same location for each image in Figure 6.

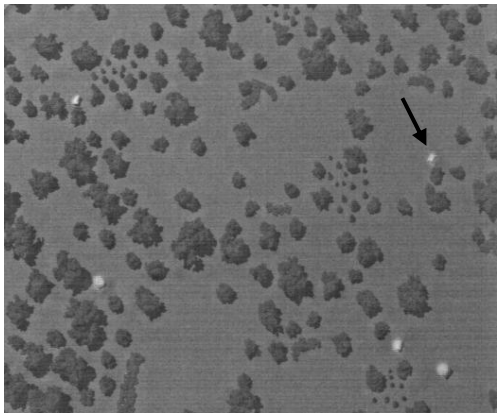
Averaging



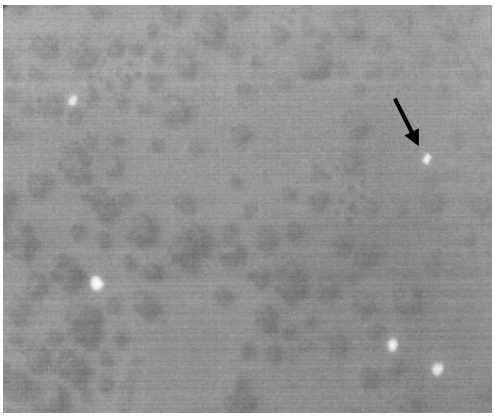
Contrast



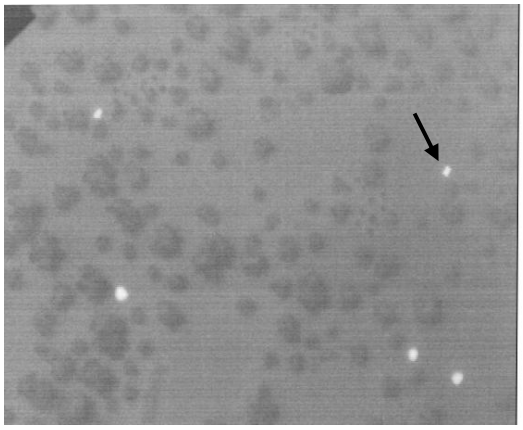
DWT



PCA



Thermal Unfused



Visible Unfused

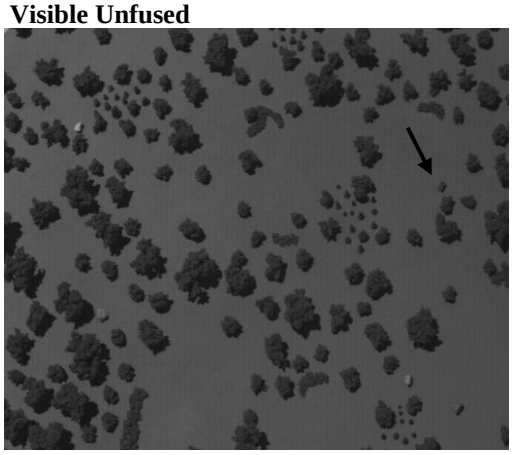


Figure 6: The six different algorithm conditions. The arrow in each image highlights the target.

5.0 RESULTS

The dependent measures of response time and percent error were measured for each trial. Including all four observers, there were 5088 trials. Of these, there were 480 catch trials. There were 163 catch trials (15%) in which the observers thought they saw a target. These trials were not used for any further analyses.

Across all observers there were 4,608 trials in which a target was present. Of these, there were 311 trials (7%) where the observer decided there was no target (before 30 seconds), and 422 trials (9%) where the observer pushed the button for the wrong quadrant. These two categories of error added up to a total error of 16%.

Both incorrect responses and timed-out responses were combined into percent error. For the 733 incorrect trials in which observers said there was no target or identified the wrong quadrant as containing the target, there is no meaningful response time for finding a target. For response time, a comparable measure of central tendency for each algorithm condition was determined within each observer using the following steps. The first step was to identify which of the 192 combinations of 16 rotation angles x 3 camera tilts x 4 quadrants contained a response time for each algorithm condition. That is, the combinations in which none of the 6 algorithm conditions had an incorrect trial. Then, for those combinations having a response time for each algorithm condition, the mean response time across these combinations was determined for each algorithm condition.

Repeated measures analyses of variance were performed with mean response time and percent error as the dependent variables. The factor was algorithm condition. *F*-tests showed a significant difference among the algorithm conditions for mean response time $\{F(5,42) = 14.94,\ p = 0.0001\}$ and percent error $\{F(5,42) = 16.11,\ p = 0.0001\}$. Post hoc paired comparisons used the Least Significant Difference (LSD) procedure with a 0.05 per comparison error level.

Figure 7 contains the results of the paired comparisons with algorithm conditions for the measure of mean response time. The results are sorted by increasing response time. The whiskers represent the least significant difference value. This value is the mean difference between a pair of algorithm conditions that would have a *p*-value of 0.05.

Figure 8 contains the results of the paired comparisons with algorithm conditions for the measure of mean percent error. The results are sorted by increasing percent error. The lower panel shows the percent error for each algorithm condition. The whiskers represent the least significant difference value (as described above).

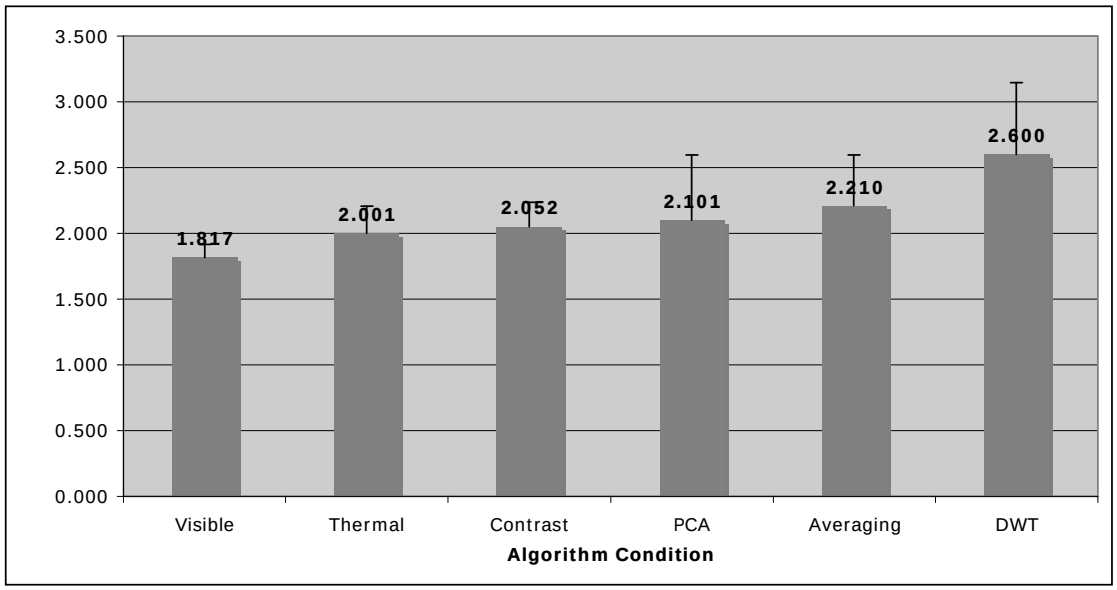


Figure 7: Mean response time (across observers) by algorithm condition.

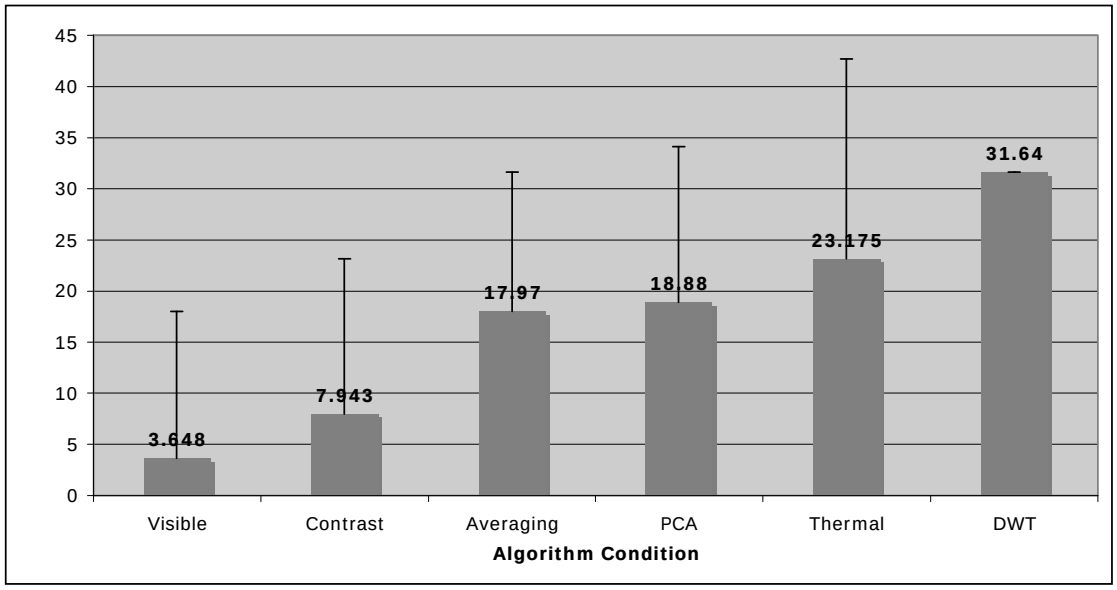


Figure 8. Mean percent error (across observers) by algorithm condition.

6.0 DISCUSSION / CONCLUSIONS

Figure 6 summarizes the results of the paired comparisons, sorted by increasing response time. As you can see, the visible unfused and thermal unfused conditions had the smallest response time. All of the fused image conditions had longer response times. The contrast fusion algorithm however was not significantly different in response time from the visible unfused or thermal unfused condition. The response time in all the other algorithm conditions was significantly different from the response time in both the Visible condition and the Thermal condition. The response time for the Contrast algorithm was not significantly different from the PCA algorithm, but was significantly different from all the other algorithm conditions.

Figure 7 summarizes the results of the paired comparisons, sorted by increasing percent error. In terms of percent error, the Visible and Contrast conditions were not significantly different from each other. However, both were significantly different from the other algorithm conditions in percent error. One can see that, in terms of percent error, both the Thermal condition and DWT condition had quite high error rates.

The objective of this work was to use the assessment methodology developed in Neriani et al.[1] to assess the degree of visual performance enhancement provided by four different fusion algorithms. Our original hypothesis was that response time and percent error would be improved by the use of fusion algorithms in a visual search task, as compared to the performance in the Visible unfused and Thermal unfused conditions. This hypothesis was not supported by the data. Based on the results discussed above, it would appear that there was no significant enhancement achieved by the fusion algorithms, at least in terms of response time. The PCA, Averaging, and DWT algorithms were all significantly different in terms of their response times from the Visible and Thermal conditions. Although, instead of having significantly shorter response times, as was expected, they all had significantly longer response times. Additionally, the PCA, Averaging, and DWT algorithms performed significantly worse than the Visible condition in terms of percent error.

It was quite unexpected that none of the four algorithms tested in this experiment showed any significant improvement in response time over the Visible and Thermal conditions. It has been a long held assumption that when spectral bands are fused to create a multi-spectral image, the result of this fusion should be both faster response time and lower percent error in a search task. The results of this study seem to suggest that this assumption may not be true in all scenarios. An alternative explanation for these results may be that the Visible image was so clear that the observers had little trouble finding the target. Therefore, there would be no need for the thermal information and thus, no benefit seen when compared to the fused conditions. In conclusion, we will further investigate the enhancing effects of fusion algorithms but will increase the difficulty of the task by making the Visible and Thermal conditions combined in a dual search task and compare performance with these images to fused conditions.

7.0 REFERENCES

K.E. Neriani, T.J. Herbranson, A.R. Pinkus, C.M. Task, H.L. Task, "Visual performance-based enhancement methodology: an investigation of three Retinex algorithms." in *Enhanced and Synthetic Vision,* J.G. Verly, ed., *Proc. SPIE 5802,* (March 2005).

S. Senthil Kumar and S. Muttan, "PCA based image fusion", in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII*, S.S. Shen and P. E. Lewis, eds., *Proc. SPIE 6233*, (April 2006).

O. Rockinger, "Image sequence fusion using a shift-invariant wavelet transform", in *Proc. IEEE Intl. Conference on Image Processing*, 288- 291 (1997).

T. Peli, E. Peli, K. Ellis and R. Stahl. "Multi-spectral image fusion for visual display", in *Sensor Fusion: Architectures, Algorithms and Applications III, Proc. SPIE 3719,* (April 1999).

8.0 APPENDIX: ANNOTATED BIBLIOGRAPHY OF SELECT IMAGE ENHANCING ALGORITHMS

Wang, Qiang and Shen, Yi (2004). The effects of fusion structures on image fusion performances. *Proceedings of IEEE 21st Conference on Instrumentation and Measurement Technology*, 1, 468 – 471.

This paper discusses the effects that different structures within a fusion process can have on the resulting fused image. The authors classify fusion structure as three entities: hierarchical fusion structure, overall fusion structure, and arbitrary fusion structure.

Hierarchical fusion structure refers to the fusion of two (and only two) source images in a predefined order. Overall fusion structures can fuse multiple images (2 and more) into one image. Most applications use both of these (hierarchical and overall) together, which is termed the arbitrary fusion structure.

This paper looks at the structures within a wavelet-based method of image fusion. When looking at a wavelet-based fusion method, the input to feature extraction and weight determination functions are different with the use of hierarchical vs. overall fusion structures. This leads to a different final fused image, as seen in their results section. However, when the authors deliberately design the feature extraction function, they get a final fused image that is the same for both fusion structures (hierarchical and overall). The authors do state, however, that the situations for which this would be applicable rarely happen.

Kwon, Oh-Kyu and Kong, Seong G. (2005). Multiscale fusion of visual and thermal images for robust face recognition. *Proceedings of IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety*, 1, 112 – 116.

This paper evaluates a discrete wavelet transform algorithm (DWT) for the fusion of visual and thermal images used for face recognition. One difficulty often encountered in the fusion of images taken of human faces is the treatment of eyeglasses. When eyeglasses are present in an image, the thermal image fails to provide useful information around the eyes since glass blocks a large portion of thermal energy. The algorithm evaluated attempts to correct for this problem.

The important points addressed by the algorithm are broken down into two "rules." Fusion Rule I finds the approximation component of the fused image by getting a weighted average of the approximate coefficients in the visual and thermal images. In the eyeglass region (if applicable) Fusion Rule I would use more visual information to enhance the visual quality of the fused image. Fusion Rule II combines all the detail components in the DWT decomposition other than the approximate component of the visual and thermal images (this is dealt with in Rule I). It uses the integration rule of selecting dominant values in the high frequency domain, which tends to preserve the salient features in the fused image.

Performance of their algorithm was compared with that of the average, Laplacian Pyramid and DWT-max fusion methods. Performance measures included both visual quality and entropy. The

authors found that the proposed method created a more visually pleasing fused image, being less sensitive to illumination changes and had better detail in the eyeglass region of the image. Additionally, the proposed method showed higher average entropy than the other methods tested.

Mitianoudis, Nikolaos and Stathaki, Tania (2006). Adaptive image fusion using ICA bases. ***Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing,*** **2, 829 – 832.**

In this paper the authors propose a fusion method based on using bases trained using Independent Component Analysis (ICA) on similar content images as analysis tools. The main motivation for this was to use bases that can fit arbitrarily on the object types that are to be fused. The authors believe that this framework can outperform generic analysis tools such as wavelet analysis.

One can train analysis bases using ICA. The training procedure needs to be completed only once, as the estimated transform can be used for fusing similar content images. A number of N x N patches (usually about 10,000) are selected from similar content training images. Principle components analysis (PCA) is performed on the selected patches and the most important bases are selected. Then, the ICA update rule is iterated for a chosen L x L neighborhood until convergence.

Once the ICA transform has been estimated, one can perform image fusion using the ICA bases. Every possible N x N patch is isolated from each image and is rearranged to form a vector. Each of the input vectors is then transformed to the ICA domain. The image patches are averaged in the same order that they are selected during the analysis step.

The authors tested their fusion algorithm using some surveillance images from three different sensors: an IR camera, a micro LW camera and a CCD camera. The authors conclude that the images using ICA bases with adaptive schemes (using a weighted vector with Laplacian priors) have increased perceived image quality.

Dawei, Zhao and Fang, Zi (2007). A new improved hierarchical model of image fusion. ***Proceedings of the 8th International Conference on Electronic Measurement and Instruments,*** **1, 2-853 – 2-857.**

This paper is a review of the basic concepts in image fusion with a description of a hierarchical model of how image fusion should proceed. The authors first review the justification for the need for image fusion. It is asserted that fusion from multiple image sources can improve decision making by providing more useful information in the fused image. The authors next review the main principles of image fusion which they state are the principles of redundancy, complementarity, time-limit, and low cost.

Next, the authors discuss the commonly referred to levels of fusion: pixel level, feature level, and decision level. This leads into their hierarchical model of image fusion. The model starts with preprocessing and includes re-sampling of the image and spatial plus temporal registration. The first level of the model is pixel fusion. Pixel-level fusion is operated in the phase of image preprocessing. This is broken into signal-level fusion and image point-level fusion. The second level of the model is feature fusion. Feature-level fusion is done in the course of image feature extraction. This prepares for decision-level fusion. At the feature level, all useful image features are extracted. These may include edges, shapes, profiles, angles, textures, similar lighting areas,

and similar depth of focus. The final level in the hierarchy is decision fusion. At this level the features obtained at the previous level are used to be able to classify and identify objects in the image and then decisions are made based on this information. The model is not applied to actual images and is not evaluated in the paper.

Singh, Harpreet; Raj, Jyoti; Kaur, Gulsheen and Meitzler, Thomas (2004). Image fusion using fuzzy logic and applications. *Proceedings of IEEE International Conference on Fuzzy Systems*, 1, 337 – 340.

This paper investigates both a fuzzy logic and neuro-fuzzy logic approach to image fusion. The authors start with a description of the fuzzy logic method. This method is to be applied for pixel-level fusion. The authors use the Fuzzy Inference System (FIS) editor of the Fuzzy logic toolbox found in Matlab. The code given in the paper outlines how the two input images are first converted into column form. Then a Fuzzy (fis) file is made (has two input images). The number and type of membership functions are then decided for both the input images by tuning the membership functions. Input images in antecedent are resolved to a degree of membership ranging from 0 to 255. There are then rules made for the input images which resolve the antecedents to a single number from 0 to 255. For each column, fuzzification is applied using the rules described above on each pixel value of the input image. This gives a fuzzy set represented by a membership function. The end result should be an output image in column form. The fused image should then be converted back to matrix form in order to be displayed.

The Neuro-fuzzy approach is similar to the fuzzy logic approach. The code given in the paper outlines how the two input images are first converted into column form. Then, training data is established in the form of a matrix with three columns, with values between 0 to 255. Next, a matrix is formed of data that is to be "checked". This is comprised of the pixels from the two input images in column format. For the training of the algorithm, the FIS structure is needed (generated by the *genfis1* command in Matlab) with training data, number and type of membership functions as the input. The *anfis* command is used to start training. Fuzzification is applied as above with the data to be "checked" and the trained data as inputs. The output fused image is then converted to matrix form for display purposes.

Both fusion processes were tested using images from both the medical domain and remote sensing domain. Entropy (Shannon's) and variance were calculated for both methods of fusion.

Ranjan, Rahul; Singh, Harpreet; Meitzler, Thomas and Gerhart, Grant R. (2005). Iterative image fusion technique using fuzzy and neuro-fuzzy logic and applications. *Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society*, 1, 706 – 710.

This method is almost entirely like the method described above in 3.5, "Image Fusion using Fuzzy Logic and Applications". However, an iterative technique is added to the method. In the case described in the paper above, when two or more images are given as inputs to the fuzzy logic or neuro-fuzzy logic method, they each have equal share pixel-wise in the final fused image. In the iterative approach that is described, priority is given to some of the images and the prioritized image is fused more than once for the final output image.

The algorithm adopts a process in which there are N number of images given as input. These images are each given a level of priority with some index of priority. This index of priority for an

image decides how many times an image has to be fused for the final output image. The authors state that this should result with a better image at two to three iterations. Matlab code is provided for both an iterative fuzzy logic approach for image fusion and an iterative neuro-fuzzy approach for image fusion. This code is mostly identical to that described above with the exception of the prioritized iterative approach added in.

Wang, H.; Peng, J. and Wu, W. (2002). Fusion algorithm for multisensor images based on discrete multiwavelet transform. *Proceedings of IEE International Conference on Vision, Image and Signal Processing*, 149(5), 283 – 289.

This paper reviews the concept of multiwavelets and describes the use of the discrete multiwavelet transform (DMWT) for image fusion. The authors present a novel algorithm for fusion. This algorithm is then applied to test images. Results of this testing are detailed.

Multiwavelets are extensions from scalar wavelets. The authors posit that they have several advantages over traditional scalar wavelets. Multiwavelets have short support, orthoganality, symmetry and a high number of vanishing moments. A scalar wavelet system cannot have all of these properties at the same time.

The first step in the image fusion algorithm described is that multiwavelet processing and decomposition of each input source image is computed at different levels. Each source image is broken down into sub-bands (sub-images). The pixels of the sub-images consist of corresponding multiwavelet decomposition coefficients. At each level, there are sixteen sub-images that can be divided into four blocks. The low-low sub-bands block shows an image's approximate part. The low-high, high-low and high-high sub-bands blocks show detail parts of the image in horizontal, vertical, and diagonal directions, respectively. A pyramid is formed for the composite image by selecting multiwavelet decomposition coefficients from the source image pyramids. In the proposed fusion scheme, the authors present a new area-based fusion rule to combine source sub-images and to form the pyramid for the composite image. Coefficients are selected between two source images' corresponding sub-bands to form the coefficients of composite sub-bands. These selected coefficients must represent the salient features in the sub-bands of the source image. The sub-bands are convolved with a feature extracting operator and a pixel is selected with a large output to the corresponding coefficient of the composite sub-bands. The fused image is then constructed by reconstructing and post filtering the combined coefficients.

Image tested was performed on two SPOT pictures of the same area. These were treated as source images. The first was a panchromatic mode image and the second is a multispectral mode image (near infrared). The images were fused with multiple methods (average, gradient pyramid, and proposed DMWT algorithm). The authors state that the best image fusion result is obtained by applying the proposed DMWT algorithm. The fused image shows that features from the source images were well preserved and properly enhanced.

Hariharan, Harishwaran; Koschan, Andreas; Abidi, Besma; Gribok, Andrei; and Abidi, Mongi (2006). Fusion of visible and infrared images using empirical mode decomposition to improve face recognition. *Proceedings of IEEE International Conference on Image Processing*, 1, 2049 – 2052.

The authors are proposing a new image fusion technique using Empirical Mode Decomposition (EMD). The paper specifically focuses on image fusion for the purpose of facial recognition, however, the algorithm itself could be generalized for other applications.

The EMD works by decomposing non-linear non-stationary signals into Intrinsic Mode Functions (IMFs). The images are decomposed from their different imaging modalities (thermal image, visible image, etc.) into IMFs. Then fusion is performed at the decomposition level and the fused IMFs are reconstructed to form the fused image. The authors tested the effectiveness of their fusion technique by looking at the Cumulative Match Characteristics (CMCs) between the test images and what was in their full gallery of images (of both visible and infrared raw images). They additionally compared results of their algorithm against results obtained using averaging, principle component analysis (PCA) fusion, and a wavelet based fusion technique.

Hu, Liangmei; Gao, Jun; He, Kefeng; and Xie, Zhao (2005). Image fusion using D-S Evidence theory and ANOVA method. ***Proceedings of the 2005 IEEE International Conference of Information Acquisition,*** **1, 427 – 431.**

This paper is a summary of a fusion method that combines an Analysis of Variance (ANOVA) method for image fusion with the Dempster-Shafer (D-S) Theory to detect weak edges. The authors propose that this coupling can be used to alleviate the sensitivity in threshold selection of the ANOVA method.

The ANOVA method is an edge detector method that relies on the value of a contrast function in each of the directional masks of an image to judge whether there is an edge or not. Normally, one uses the horizontal, vertical and two diagonal directions to stand for all the possible edge directions in a small mask. Whether an edge is detected or not depends on the pixel value distribution pattern in the directional mask that is tested and the value of the threshold coefficient that is chosen. The ANOVA method is sensitive to the value chosen for its threshold coefficient, *k*. If *k* is too high, a weak edge may be overlooked. However, if *k* is too low, noisy pixels may be accepted as edge pixels. To deal with this the D-S method is applied.

D-S Evidence Theory is a statistical-based data fusion classification algorithm.

Jiang, Lijun; Tian, Feng; Shen, Lim Ee; Wu, Shiqian; Yao, Susu; Lu, Zhongkang; and Xu, Lijun (2004). Perceptual-based fusion of IR and visual images for human detection. ***Proceedings of the 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing,*** **1, 514 – 517.**

The image fusion algorithm that the authors of this paper propose is a combination of the Multiple Scale fusion approach and the Perceptual-based fusion approach. The proposed fusion method is based on the contrast sensitivity of the human visual system.

The first step in the fusion approach is that the perceptual contrast difference, D, is computed. This difference is calculated based on the saliency of the current pixel in both the images to be fused (the visible and thermal, as is the case in this paper). Saliency is a level of prominence of a pixel relative to its neighboring pixels. Saliency is calculated at different scale sizes. The authors use scale sizes of 3, 5, 7, 9, and 11 in this paper. After computing the perceptual contrast difference, it is compared to some threshold, T. In this paper, the value of 0.25 is set for T for all fusion scales, based on testing results completed by the authors. If D is greater than the threshold, T, then the

contrast value of the pixel with the higher saliency value will be retained in the fused image. If D is less than T, the contrast value of each image pixel will have their weights computed given the equation found in the paper. Then the fused image is created by summing the weighted contrast value of each pixel from the visible image and the thermal image. Post-processing with a Laplacian transform is suggested for image enhancement in the areas of gray-level discontinuities. This should sharpen the features of the fused images.

The proposed fusion method is tested by the authors on the standard benchmark "Clock" images. These images are used to compare the proposed fusion method versus a linear averaging fusion method. The authors conclude that comparatively, the proposed method performs quite well. Additionally, the proposed fusion approach is tested with human detection in a dimly lit room. Images from both a thermal and visible camera are fused using the proposed method with various different filter combinations to remove noise. It is found that the proposed method works best in conjunction with post-processing with the average filter and Laplacian filter, since these two filters complement each other.

Peli, Tamar; Young, Mon; Knox, Robert; Ellis, Ken and Bennett, Fredrick (1999b). Feature level sensor fusion. ***Proceedings of SPIE: Sensor Fusion: Architectures, Algorithms and Applications III,*** **3719, 332 – 339.**

The authors discuss the use of two different fusion techniques in this paper. The hybrid fusion and cued fusion techniques are to be used for automatic target cuing that serves to combine features obtained from each sensor's image at the object level.

In the hybrid fusion method, the first step is to prescreen the data coming from each sensor. This is referred to as Automatic Target Cuing (ATC). This is done prior to the fusion stage.

The cued fusion method assumes that one of the sensors will be designated as the "primary" sensor by an operator. The second sensor will be used for false alarm reduction. The ATC is then performed only on the primary sensor's input data. If one of the sensors has a higher probability of detection (and/or a lower false alarm rate), it can be selected as the primary sensor. However, if the ground coverage can be segmented to regions in which one of the sensors is known to exhibit better performance, then the cued fusion can be applied locally/adaptively by switching the choice of a primary sensor. Otherwise, the cued fusion is applied both ways (each sensor as a primary) and the outputs of the cued mode are combined.

Both the hybrid fusion method and the cued fusion method use a back-end discrimination stage. This is applied to a combined feature vector to reduce false alarms. These two fusion methods were used with spectral and radar data. They were each shown to substantially reduce false alarms.

Petrovic, Vladimir and Xydeas, Costas (1999). Cross band pixel selection in multi-resolution image fusion. ***Proceedings of SPIE: Sensor Fusion: Architectures, Algorithms and Applications III,*** **3719, 319 – 326.**

This paper describes an image fusion technique using cross band pixel selection. The fusion is realized using multiresolution analysis and synthesis by Quadrature Mirror Filter (QMF) banks. The authors investigated this fusion algorithm with the aim of reducing the contrast and structural

distortion image artifacts that are produced by traditional wavelet-based, pixel-level fusion techniques.

The Quadrature Mirror Filters are a special class of Forward Infrared (FIR) filters that are used for sub-band decomposition in a tree structure that is capable of perfect signal reconstruction. Conventional QMF banks are made up of pairs of complimentary FIR filters. The QMF used in this image fusion study are half-band FIR filters.

The image fusion system aims to increase the information content of the resulting single images by selecting the most "significant features" from all of the given input images and then transferring them into the composite fused image. This is done by creating a new fused pyramid using the multiresolution QMF sub-band pyramids produced from the input images. The process occurs by having the input images filtered using the series of multiresolution QMF Analysis filter pairs. This decomposes the images into multiresolution pyramid structures. These sub-band signals are then examined via a feature selection process. Selection decisions are then made for all pyramid levels and pixel positions. To yield the fused pyramid a QMF synthesis process is applied to the fused sub-band signals to get a fused image.

The study reviews two different methods of pixel selection: a traditional QMF based decomposition system that would use an area-based sub-band selection for pyramid fusion (Scheme 1) and the cross-band selection strategy (Scheme 2) developed by the authors. The cross-band method uses a two stage process as described in the paper; pairs of input images (visible and infrared) registered images which were processed with Scheme 1 and Scheme 2. A subjective test was completed using 11 participants and 10 pairs of input images. For each of the image pairs, participants expressed a preference in favor of one of the fused images. This resulted in a normalized score, P, per input image. P was calculated by dividing the number of participants indicating a preference for each scheme by the total number of participants. The P score for Scheme 1 (the area-based selection method) was lower for all the images than the P score for Scheme 2 (cross-band selection strategy). The authors state that their proposed strategy tends to exhibit considerably fewer artifacts in terms of shadow and ringing effects and therefore provides an advantage over the area-based selection strategy.

Li, H.; Manjunath, B. S. and Mitra, S. K. (1995). Multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing*, 57(3), 235 – 245.

This paper presents an image fusion scheme based on the wavelet transform. The authors describe the basic concept of wavelet transform for fusion first. They then outline the advantages that the wavelet transform has over Laplacian Pyramid style techniques. These include the fact that the wavelet transform is the same size as the image, the wavelet transform takes into account spatial orientation selectivity and that information contained at different resolutions is unique with the wavelet transform.

The authors next review the implementation of their algorithm for the purpose of image fusion. The input images are first broken down into low-high, high-low, high-high and low-low bands. While a standard technique would be to select the larger of the two wavelet coefficients at each point to be included in the fused image, the authors implement a modified feature selection algorithm. This can help better identify features for selection, resulting in a fused image that is higher in visual quality. The algorithm operates on an area-based selection rule. The images are decomposed into a gradient

pyramid. The variance of each image patch over a 3x3 or 5x5 window is computed as an activity measure associated with the pixel at the center of the window. A binary decision map of the same size as the wavelet transform is created to record the selection results based on the feature selection rule.

The algorithm is evaluated on its performance in the next section. Various different pairs of multisensor images were evaluated, including multifocus images, MRI and PET images, Landsat TM and SPOT images, Landsat TM and Seasat SAR images, and visible and infrared images. Overall, the authors conclude that images fused using the proposed fusion method had a higher image quality rating over those fused by pixel-by-pixel averaging and a Laplacian Pyramid method.

Lindberg, Perry C.; Dasarathy, Belur V. and McCullough, Claire (1996). Multi-level fusion exploitation. *Proceedings of SPIE: Signal Processing, Sensor Fusion, and Target Recognition V,* 2755, 260 – 270.

The authors of this paper describe a project designed to exploit the use of sensor fusion at all levels, including signal, feature and decision levels. These are to be used to improve target recognition capability against Tactical Ballistic Missile (TBM) targets.

The objective of the experiments described was to develop, test, and evaluate Automatic Target Recognition (ATR) algorithms by comparing their performance using fused sensor information at the different levels of abstraction (signal, feature, decision, and various combinations of these). The authors created target recognition algorithms that fuse sensor information at each of the above levels. These algorithms were trained with simulated target data and tested with realistic target signatures. The scenario chosen was a ship platform that had two radar sensors (S-band and X-band) that can simultaneously observe TBMs targets.

The authors investigated a maximum likelihood scheme, a nearest-neighbor scheme, a piecewise sequential scheme, a neural network strategy, and two different decision fusion schemes. The authors concluded that none of the classifiers had superior performance over all the target types that were tested, but some classifiers were superior at recognizing particular targets. The test results showed the benefits of combining sensor fusion decisions at all levels to improve target recognition performance. The biggest improvement in fusion occurred when the outputs from two moderately successful classifiers were combined.

Wang, Hai-Hui; Zhang, Jian; Wang, Jun and Wang, Wei (2005). A novel method based on discrete multiple wavelet transform to multispectral image fusion. *Proceedings of SPIE: International Symposium on Multispectral Image Processing and Pattern Recognition, Image Analysis Techniques*, 6044, 60440T-1 – 60440T-8.

The authors of this paper first review the concept of multiwavelets and then describe the use of the discrete multiwavelet transform (DMWT) for image fusion. The authors present a novel fusion algorithm. This algorithm is then applied to test images. Results of this testing are detailed.

Multiwavelets are extensions from scalar wavelets. The authors posit that they have several advantages over traditional scalar wavelets. Multiwavelets have short support, orthoganality, symmetry, and a high number of vanishing moments. A scalar wavelet system cannot have all of these properties at the same time.

The first step in the image fusion algorithm described is that multiwavelet processing and decomposition of each input source image is computed at different levels. Each source image is broken down into sub-bands (sub-images). The pixels of the sub-images consist of corresponding multiwavelet decomposition coefficients. At each level there are sixteen sub-images that can be divided into four blocks. The low-low sub-bands block shows an image's approximate part. The low-high, high-low and high-high sub-bands blocks show detailed parts of the image in horizontal, vertical, and diagonal directions, respectively. A pyramid is formed for the composite image by selecting multiwavelet decomposition coefficients from the source image pyramids. In the proposed fusion scheme, the authors present a new area-based fusion rule to combine source sub-images and to form the pyramid for the composite image. Coefficients are selected between two source images' corresponding sub-bands to form the coefficients of composite sub-bands. These selected coefficients must represent the salient features in the sub-bands of the source image. The sub-bands are convolved with a feature-extracting operator and a pixel is selected with a large output to the corresponding coefficient of the composite sub-bands. The fused image is then constructed by reconstructing and post filtering the combined coefficients.

Testing of the fusion algorithm was completed on two SPOT images (a panchromatic image and a multispectral mode image). The results of the fusion method were compared against those for averaging, gradient pyramid and the standard DWT (Daubechies D4 scalar wavelet). The results show that taking an average reduces the contrast of features in the source images. The gradient pyramid gives a better result, but some of the edges are blurred. The fusion method based on DWT performs better than the method based on the gradient pyramid. The best image fusion result, however, is obtained by applying the proposed DMWT algorithm. The details were well-preserved and even enhanced.

Kumar, S. Senthil and Muttan, S. (2006). PCA based image fusion. ***Proceedings of SPIE: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII, 6233,*** **62331T-1 – 62331T-7.**

This paper reviews an image fusion method based on the concept of Principle Components Analysis (PCA). Principle Components Analysis is a statistical technique used to decrease a large set of variables to a much smaller set that still contains most of the information that was available in the larger set. By reducing the interchannel dependencies to pull out just the principle factors (or components) of the data, it becomes much easier to analyze and interpret. The basic concept behind using a PCA method to fuse images is to find the unique information in each of the spectral bands and create a new image by fusing only that non-redundant information that contributes the most to the variation in the data set.

The method is tested on images obtained from a visible camera and an infrared camera. The fused image appears more contrast enhanced than it would if another fusion technique were employed (e.g., averaging or superposition).

Dixon, Timothy D.; Noyes, Jan; Troscianko, Tom; Canga, Eduardo Fernandez; Bull, Dave; and Canagarajah, Nishan (2005). Psychophysical and metric assessment of fused images. ***Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization*****, 95, 43 – 50.**

In this study, the authors investigated the effects of image compression and image fusion on image usability. The study used a task-based method of image assessment. A signal detection paradigm was used, in which participants were required to identify the presence or absence of a target in briefly presented images followed by an energy mask. These results were compared with computational metric results.

In the first experiment, eighteen participants were shown fused images of infrared and visible light images. In each case, a target (a soldier) was present in the image or not. Participants were required to indicate whether a target was present or not upon presentation of each image. There were two independent variables to this experiment, each with three levels: the image fusion method that was used (averaging, contrast pyramid, and dual-tree wavelet transform) and JPEG2000 compression (no compression, low compression and high compression). The images were shown in a repeated measures design. Task performance and metric results are collected and compared. The images were blocked by fusion type, with compression type randomized within blocks. Experiment two was the same as experiment one, except JPEG images were substituted for JPEG2000 images. Results showed that there was a significant effect of fusion technique, but not compression for the JPEG2000 images. There was a significant effect for both fusion technique and compression for the JPEG images. Overall, the dual-tree wavelet transform had the best results for both the human target detection task and the computational metrics. The contrast pyramid was shown to underperform, particularly with the JPEG results in the human performance measure. The averaging method was seen as the least favorable, with the exception of the JPEG results in the human performance measure.

Lewis, John J.; O'Callaghan, Robert J.; Nikolov, Stavri G.; Bull, David R. and Canagarajah, Nishan (2005). Pixel- and region-based image fusion with complex wavelets. ***Information Fusion*****, 8(2), 119 – 130.**

This paper highlights several pixel-based fusion algorithms and compares their performance with a novel region-based image fusion method. Pixel-based image fusion is reviewed, including spatial image fusion, multi-resolution image fusion, pyramid transform fusion, and wavelet transform fusion.

The authors then describe a region-based fusion method that they claim holds a distinct advantage over the traditional pixel-based fusion methods described above. The advantages include intelligent fusion rules, highlighted features, reduced sensitivity to noise, and the ability to register images and complete video fusion. A dual-tree complex wavelet transform is used to segment the features of the input images, either jointly or separately, to produce a region map. Characteristics of each region are calculated and a region-based approach is used to fuse the images, region by region, in the wavelet domain.

The methods are evaluated and compared on their ability to fuse infrared and visible images of an outdoor scene and a multifocus image of an indoor scene. Three different evaluation metrics are considered: mutual information, the Xydeas and Petrovic metric, and the Piella and Heijnmans

image quality metric. The region-based image method produces images with better contrast that more accurately reflects the contrast of the original images compared with the pixel-based fusion methods.

Wang, Hai-Hui; Zhang, Jian and Wang, Wei (2005). Fusion algorithm for images data by using steerable pyramid transform. *Proceedings of IEEE International Conference on Machine Learning and Cybernetics*, 8, 5050 – 5054.

In this paper, a novel fusion algorithm is presented for multisensor images using a steerable pyramid transform that is performed at the pixel level. A match and saliency measure is used to combine the detail images. The saliency measure is based on the local energy in the detail domain. The saliency measure can be defined as the local energy of the incoming pattern within a neighborhood. The salience of a particular component is high if that pattern plays a role in representing important information in a scene, and it is low if the pattern represents unimportant information (or corrupted image data).

After the saliency computation step has been applied to the pyramids for the input images, A and B, a match measure has to be computed to combine the information carried by each pyramid. This is identified in the paper.

The authors investigated the performance of the steerable pyramid described above and the performance of a traditional Laplacian Pyramid design at fusing two multifocus images. Mutual information (MI) is used as a performance measure for both fusion methods. The results show that the fused image using the fusion algorithm based on a steerable pyramid has higher MI. The authors state that based on this measure, the steerable pyramid outperforms the Laplacian Pyramid fusion method.

Rockinger, Oliver (1997). Image sequence fusion using a shift-invariant wavelet transform. *Proceedings of IEEE International Conference on Image Processing*, 1, 288- 291.

The author describes a generic wavelet fusion scheme and investigates its shift dependency. The disadvantages of shift dependency are reviewed. The importance of using a shift-invariant wavelet fusion method versus a standard wavelet fusion method comes into play when fusing images that have unknown object locations.

In standard fusion methods, one must have fixed object locations to achieve a pleasing fused image. In the shift-invariant method, the images are decomposed such that all possible (circular) shifts of the input images are calculated. This is overly complete and redundant. In the fusion process described in this paper, the input images are decomposed into their shift-invariant wavelet representation and a fused representation is built by using an appropriate selection scheme. The author identifies two methods of selection. The method chosen for this study is the point based *choose-max* method.

In terms of evaluation, an information theoretic quality measure evaluates the temporal stability of the fused images. The method described is compared to other existing multi-resolution fusion schemes, including the Laplacian Pyramid fusion method and the discrete wavelet transform on two out-of-focus input images. The root mean square error between the ideal image and the actual

fusion result was calculated. The best results occurred for the shift-invariant wavelet transform that the author describes.

Laporterie-Dejean, F.; Latry, C. and De Boissezon, H. (2003). Evaluation of the quality of panchromatic/ multispectral fusion algorithms performed on images simulating the future Pleiades satellites. *Proceedings of the 2nd GRSS/ISPRS Joint Workshop on Data Fusion and Remote Sensing over Urban Areas*, 1, 95 – 98.

The authors review the results of a study comparing five fusion methods used for both panchromatic and multispectral images. The goal of the study was to find an existing fusion method that would be considered well-suited for the high resolution images coming from the Pleiadas-HR satellites. The satellites have one panchromatic channel (500 – 850 nm), and four multispectral channels. These are a blue channel (430 -550 nm), a green channel (490 – 610 nm), a red channel (600 – 720 nm), and a near infrared channel (750 – 950 nm). The five methods include: the ARSIS (Amlioration de la Resolution Spatiale par Injection de Structure) method, the GLP (Generalized Laplacian Pyramid) method, the LSM (Local Least Squares Modeling) method, the LMVM (Local mean and Variance Matching) method, and the CNES (Centre National d'Etudes Spatiales – French Space Agency) method.

In the ARSIS method, the panchromatic band is decomposed by a wavelet transform. Then the inverse wavelet transform is performed with the coefficients calculated and using one of the multispectral channels. This process is iterated for each of the multispectral channels. In the GLP method, the panchromatic channel is decomposed through a pyramid algorithm process. The inverse transform of the Laplacian is then performed with one multispectral channel and the high frequencies that are extracted during the decomposition. This process is also iterated for each multispectral channel. In the LSM method a linear regression is performed between the degraded panchromatic channel and one multispectral channel. Then the panchromatic channel and the regression coefficients are used to synthesize a high resolution multispectral channel. This process is iterated for each of the multispectral channels. In the LMVM method, the principle is to change the space representation of the multispectral images and then calculate an intensity image. The inverse transform is calculated using the panchromatic image instead of the intensity image. The CNES method is based on the principles of human vision. Three multispectral channels are transformed from the RGB color space to the IHS color space. The inverse transform is performed using the intensity image modulated by the panchromatic band together with the hue and saturation images. If the aim of fusion is to create a natural color product, then the three channels that should be used are the red, green and blue wavelengths. Otherwise, if the aim of the fusion is to generate a false color composition, the three channels that should be used are the green, red, and near infrared wavelengths.

The fusion methods were evaluated as to their appropriateness using both quantitative and qualitative methods. Quantitative measures included mean and standard deviation of the difference between reference and merged images, relative error from the mean, standard deviation and range, and mean and standard deviation of the contrast image and correlation.

According to the authors, in terms of a quantitative evaluation, the CNES and LSM methods seem best for color indicators, while the CNES, ARSIS and GLP are best when considering single-band indicators. In terms of a qualitative evaluation, the experts overall use of the evaluation, the GLP and CNES methods were most appreciated. The ARSIS and LMVM methods were seen as poor,

and the LSM method was seen as generally good but created some unexpected artifacts, and thus, was penalized.

Bender, Edward J.; Reese, Colin E.; and van der Wal, Gooitzen S. (2003). Comparison of additive image fusion vs. feature-level image fusion techniques for enhanced night driving. ***Proceedings of SPIE: Low-Light-Level and Real-time Imaging Systems, Components, and Applications,*** **4796, 140 – 151.**

The authors review a feature-level fusion methodology for use with their Head-tracked Vision System (HTVS) program. The HTVS is a driving system for wheeled and tracked military vehicles. The system uses dual-waveband sensors that are directed in a more natural head-slewed imaging mode. The HTVS consists of thermal and image-intensified TV sensors, a high-speed gimbal, a head-mounted display, and a head tracker.

Prior to this research being conducted, the researchers had focused on the benefits that were achieved when the images from the system were combined using an additive (sensor A + sensor B) image fusion method. These benefits were assessed in terms of the enhancement to an operator's overall driving performance. The additive fusion method uses a single (operator adjustable) fractional weighting for all the features of each sensor's image.

The new feature-level fusion method that is reviewed is a multi-resolution pyramid technique. This method uses digital processing techniques to select, at each image point, only the sensor that has the strongest features. Only these strong features are used to reconstruct the fused video image. The selection process is performed simultaneously at multiple scales of the image. These scales are combined to form the reconstructed fused image.

The authors evaluate both the additive fusion method and the proposed feature-level fusion method on a variety of images using an IITV sensor and a thermal sensor. The authors conclude that the feature-level fused images tend to show clearly improved feature resolution and improved contrast over the additive fused images. However, they do note that the feature-level fusion approach does tend to result in more noise preserved from the IITV image. The authors suggest that this can be reduced by using a feature-level-tuning technique.

Peli, Tamar; Peli, Eli; Ellis, Ken and Stahl, Robert (1999a). Multi-spectral image fusion for visual display. ***Proceedings of SPIE: Sensor Fusion: Architectures, Algorithms and Applications III,*** **3719, 359 – 368.**

This paper describes a contrast-based monochromatic fusion process. The fusion process is aimed for on board real-time application and it is based on practical and computationally efficient image processing components. The process maximizes the information content in the fused image, while retaining visual cues that are essential for navigation/piloting tasks.

The multispectral fusion process described in this paper has three key attributes: 1. A scale-by-scale fusion using oriented filters, 2. A fusion decision based on the contrast in each scale and 3. A preference for the visible band for at least larger scale sizes (to preserve shape from shading contrast). The input images are filtered at four orientations (0°, 45°, 90°, and 135°). The basic fusion process compares a calculated contrast measure for each pixel, at each scale and orientation,

and selects the spectral band that should dominate in the fused image. An important consideration is that even when the combination rule is a binary selection, the fused image may have a combination of pixel values taken from the two components at various scales since it is taken at each scale.

The fusion concept was tested with imagery from image intensifiers and infrared sensors. The fused image maintains the shape from shading of the visual band for large objects (hills and valleys). Small features are taken mainly from the infrared sensor. The authors conclude that this fused image will result in optimal human performance in a piloting task.

3.0 ABBREVIATIONS AND ACRONYMS

ANOVA - Analysis of Variance
ARSIS - Amlioration de la Resolution Spatiale par Injection de Structure
ATC - Automatic Target Cuing
ATR - Automatic Target Recognition
CCD - Charge-coupled Device
CMCs - Cumulative Match Characteristics
CNES - Centre National d'Etudes Spatiales (French Space Agency)
DMWT - Discrete Multi-wavelet Transform
D-S - Dempster-Shafer
DWT - Discrete Wavelet Transform
DWT-max - Discrete Wavelet Transform with Maximum coefficients selected
EMD - Empirical Mode Decomposition
FIR - Forward Infrared
FIS - Fuzzy Inference System
GLP - Generalized Laplacian Pyramid
GHM - Geronimo, Hardin, and Massopust multiscaling function
HTVS - Head-tracked Vision System
ICA - Independent Component Analysis
IITV - Image-intensified Television
IMFs - Intrinsic Mode Functions
IR - Infrared
JPEG - Joint Photographic Experts Group
JPEG2000 - Joint Photographic Experts Group standard 2000
Landsat TM - Landsat Thematic Mapper
LMVM - Local Mean and Variance Matching
LSD - Least Significant Difference
LSM - Local Least Squares Modeling
LW - Long wave
MI - Mutual Information
MRI - Magnetic Resonance Imaging
MSR - Multiscale Retinex
PCA - Principle Component Analysis
PET - Positron Emission Tomography
QMF - Quadrature Mirror Filter
Seasat SAR - Seasat Synthetic Aperture Radar
SNR - Signal-to-noise ratio
SPOT - Satellite Pour l'Observation de la Terre
SSR - Single-scale Retinex
TBM - Tactical Ballistic Missile
TV - Television